

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: <b>C12N 15/54, 9/12, C12Q 1/68</b>		A1	(11) International Publication Number: <b>WO 96/38568</b> (43) International Publication Date: <b>5 December 1996 (05.12.96)</b>
(21) International Application Number: <b>PCT/US96/06906</b>		(81) Designated States: CA, JP.	
(22) International Filing Date: <b>14 May 1996 (14.05.96)</b>		Published <i>With international search report.</i>	
(30) Priority Data: <b>08/455,686 31 May 1995 (31.05.95) US</b>			
(71) Applicant: <b>AMERSHAM LIFE SCIENCE, INC. [US/US]; 26111 Miles Road, Cleveland, OH 44128 (US).</b>			
(72) Inventors: <b>DAVIS, Maria; 12080 Nathaniel Lane, Twinsburg, OH 44087 (US). MOFFETT, R., Bruce; 3270 Kenmore Road, Shaker Heights, OH 44122 (US). FULLER, Carl, W.; 3397 E. Monmouth, Cleveland Heights, OH 44118 (US).</b>			
(74) Agents: <b>WARBURG, Richard J. et al.; Lyon &amp; Lyon, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071-2066 (US).</b>			
(54) Title: <b>THERMOSTABLE DNA POLYMERASES</b>			
(57) Abstract			
<p>An enzymatically active DNA polymerase having between 540 and 582 amino acids having a tyrosine at a position equivalent to position 667 of Taq DNA polymerase, wherein said polymerase lacks 5' to 3' exonuclease activity, and wherein said polymerase has at least 95 % homology in its amino acid sequence to the DNA polymerase of <i>S.(i)Thermus aquaticus</i>, <i>S.(i)Thermus flavus</i> or <i>S.(i)Thermus thermophilus</i>, and wherein said polymerase forms a single polypeptide band on an SDS PAGE.</p>			

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

DESCRIPTIONThermostable DNA polymerasesBackground of the Invention

The present invention relates to novel thermostable DNA polymerases, the genes and vectors encoding them and their use in DNA sequencing.

5 US Patents 4,889,818 and 5,079,352 describe the isolation and expression of a DNA polymerase known as Taq DNA Polymerase (hereinafter referred to as Taq). It is reported that amino-terminal deletions wherein approximately one-third of the coding sequence is absent  
10 have resulted in producing a gene product that is quite active in polymerase assays. Taq is described as being of use in PCR (polymerase chain reaction).

US Patent No. 5,075,216 describes the use of Taq in DNA sequencing.

15 International patent application WO 92/06/06188 describes a DNA polymerase having an identical amino acid sequence to Taq except that it lacks the N-terminal 235 amino acids of Taq and its use in sequencing. This DNA polymerase is known as  $\Delta$  Taq.

20 US Patent 4,795,699 describes the use of T7 type DNA polymerases (T7) in DNA sequencing. These are of great use in DNA sequencing in that they incorporate dideoxy nucleoside triphosphates (NTPs) with an efficiency comparable to the incorporation of deoxy  
25 NTPs; other polymerases incorporate dideoxy NTPs far less efficiently which requires comparatively large

quantities of these to be present in sequencing reactions.

At the DOE Contractor-Grantee Workshop (Nov. 13-17, 1994, Santa Fe) and the I. Robert Lehman Symposium (Nov 5 11-14, 1994, Sonoma), Prof. S. Tabor identified a site in DNA polymerases that can be modified to incorporate dideoxy NTPs more efficiently. He reported that the presence or absence of a single hydroxy group (tyrosine vs. phenylalanine) at a highly conserved position on *E. coli*, DNA Polymerase I, T7, and Taq makes more than a 1000-fold difference in their ability to discriminate against dideoxy NTPs. (See also European Patent Application 94203433.1 published May 31, 1995, Publication No. 0 655 506 A1 and hereby incorporated by 10 reference herein.)

#### Summary of the Invention

The present invention provides a DNA polymerase having an amino acid sequence differentiated from Taq in that it lacks the N-terminal 272 amino acids and has the 20 phenylalanine at position 667 (of native Taq) replaced by tyrosine. Preferably, the DNA polymerase has methionine at position 1 (equivalent to position 272 of Taq) (hereinafter referred to as FY2) The full DNA sequence is given as Fig 1 (SEQ. ID. NO. 1). Included 25 within the scope of the present invention are DNA polymerases having substantially identical amino acid sequences to the above which retain thermostability and efficient incorporation of dideoxy NTPs.

By a substantially identical amino acid sequence is meant a sequence which contains 540 to 582 amino acids that may have conservative amino acid changes compared with Tag which do not significantly influence

5 thermostability or nucleotide incorporation, i.e. other than the phenylalanine to tyrosine conversion. Such changes include substitution of like charged amino acids for one another, or amino acids with small side chains for other small side chains, e.g., ala for val. More

10 drastic changes may be introduced at noncritical regions where little or no effect on polymerase activity is observed by such a change.

The invention also features DNA polymerases that lack between 251 and 293 (preferably 271 or 272) of the

15 N-terminal amino acids of *Thermus flavus* (Tfl) and have the phenylalanine at position 666 (of native Tfl) replaced by tyrosine; and those that lack between 253 and 295 (preferably 274) of the N-terminal amino acids of *Thermus thermophilus* (Tth) and have the phenylalanine

20 at position 669 (of native Tth) replaced by tyrosine.

By efficient incorporation of dideoxy NTPs is meant the ability of a polymerase to show little, if any, discrimination in the incorporation of ddNTPs when compared with dNTPs. Suitably efficient discrimination

25 is less than 1:10 and preferably less than 1:5. Such discrimination can be measured by procedures known in the art.

One preferred substantially identical amino acid sequence to that given above is that which contains 562

30 amino acids having methionine at position 1 and alanine

at position 2 (corresponding to positions 271 and 272 of native Taq) (hereinafter referred to as FY3). A full DNA sequence is given as Fig. 2. This is a preferred DNA polymerase for expression by a gene of the present

5 invention.

The purified DNA polymerases FY2 and FY3 both give a single polypeptide band on SDS polyacrylamide gels, unlike  $\Delta$  Taq, having either a phenylalanine or tyrosine at position 667 which forms several polypeptide bands of

10 similar size on SDS polyacrylamide gels.

A second preferred substantially identical amino acid sequence is that which lacks 274 of the N-terminal amino acids of Thermus thermophilus having methionine at position 1, and the phenylalanine to tyrosine mutation at position 396 (corresponding to position 669 of native Tth) (hereinafter referred to as FY4). A full DNA sequence is given as Fig. 5 (SEQ. ID. NO. 14).

The present invention also provides a gene encoding a DNA polymerase of the present invention. In order to assist in the expression of the DNA polymerase activity, the modified gene preferably codes for a methionine residue at position 1 of the new DNA polymerase. In addition, in one preferred embodiment of the invention, the modified gene also codes for an alanine at position

25 2 (corresponding to position 272 of native Taq).

In a further aspect, the present invention provides a vector containing the gene encoding the DNA polymerase activity of the present invention, e.g., encoding an amino acid sequence differentiated from native Taq in

30 that it lacks the N-terminal 272 amino acids and has

phenylalanine at position 396 (equivalent to position 667 of Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

In a yet further aspect, the present invention

5 provides a host cell comprising a vector containing the gene encoding the DNA polymerase activity of the present invention, *e.g.*, encoding an amino acid sequence differentiated from native Taq in that it lacks the N-terminal 272 amino acids and has phenylalanine at

10 position 396 (equivalent to position 667 of native Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

The DNA polymerases of the present invention are preferably in a purified form. By purified form is

15 meant that the DNA polymerase is isolated from a majority of host cell proteins normally associated with it; preferably the polymerase is at least 10% (w/w) of the protein of a preparation, even more preferably it is provided as a homogeneous preparation, *e.g.*, a

20 homogeneous solution. Preferably the DNA polymerase is a single polypeptide on an SDS polyacrylamide gel.

The DNA polymerases of the present invention are suitably used in sequencing, preferably in combination with a pyrophosphatase. Accordingly, the present

25 invention provides a composition which comprises a DNA polymerase of the present invention in combination with a pyrophosphatase, preferably a thermostable pyrophosphatase such as *Thermoplasma acidophilum* pyrophosphatase. (Schafer, G. and Richter, O.H. (1992)

30 Eur. J. Biochem. 209, 351-355).

The DNA polymerases of the present invention can be constructed using standard techniques. By way of example, mutagenic PCR primers can be designed to incorporate the desired Phe to Tyr amino acid change (FY 5 mutation) in one primer. In our hands these primers also carried restriction sites that are found internally in the sequence of the Taq polymerase gene clone of Delta Taq, pWB253, which was used by us as template DNA. However, the same PCR product can be generated with this 10 primer pair from any clone of Taq or with genomic DNA isolated directly from *Thermus aquaticus*. The PCR product encoding only part of the gene is then digested with the appropriate restriction enzymes and used as a replacement sequence for the clone of Delta Taq digested 15 with the same restriction enzymes. In our hands the resulting plasmid was designated pWB253Y. The presence of the mutation can be verified by DNA sequencing of the amplified region of the gene.

Further primers can be prepared that encode for a 20 methionine residue at the N-terminus that is not found at the corresponding position of Taq, the sequence continuing with amino acid residue 273. These primers can be used with a suitable plasmid, e.g., pWB253Y DNA, as a template for amplification and the amplified gene 25 inserted into a vector, e.g., pRE2, to create a gene, e.g., pRE273Y, encoding the polymerase (FY2). The entire gene can be verified by DNA sequencing.

Improved expression of the DNA polymerases of the present invention in the pRE2 expression vector was 30 obtained by creating further genes, pREFY2pref (encoding

a protein identical to FY2) and pREFY3 encoding FY3. A mutagenic PCR primer was used to introduce silent codon changes (i.e., the amino acid encoded is not changed) at the amino terminus of the protein which did not affect 5 the sequence of the polypeptide. These changes led to increased production of FY2 polymerase. FY3 was designed to promote increased translation efficiency *in vivo*. In addition to the silent codon changes introduced in pREFY2pref, a GCT codon was added in the 10 second position (SEQ. ID. NO. 2), as occurs frequently in strongly expressed genes in *E. coli*. This adds an amino acid to the sequence of FY2, and hence the protein was given its own designation FY3. Both constructs produce more enzyme than pRE273Y.

15 Silent codon changes such as the following increase protein production in *E. coli*:

substitution of the codon GAG for GAA;

substitution of the codon AGG, AGA, CGG or CGA for CGT or CGC;

20 substitution of the codon CTT, CTC, CTA, TTG or TTA for CTG;

substitution of the codon ATA for ATT or ATC;

substitution of the codon GGG or GGA for GGT or GGC.

The present invention also provides a method for determining the nucleotide base sequence of a DNA 25 molecule. The method includes providing a DNA molecule annealed with a primer molecule able to hybridize to the DNA molecule; and incubating the annealed molecules in a vessel containing at least one deoxynucleotide triphosphate, and a DNA polymerase of the present 30 invention. Also provided is at least one DNA synthesis

terminating agent which terminates DNA synthesis at a specific nucleotide base. The method further includes separating the DNA products of the incubating reaction according to size, whereby at least a part of the 5 nucleotide base sequence of the DNA molecule can be determined.

In preferred embodiments, the sequencing is performed at a temperature above 50°C, 60°C, or 70°C.

In other preferred embodiments, the DNA polymerase 10 has less than 1000, 250, 100, 50, 10 or even 2 units of exonuclease activity per mg of polymerase (measured by standard procedure, see below) and is able to utilize primers having only 4, 6 or 10 bases; and the concentration of all four deoxynucleoside triphosphates 15 at the start of the incubating step is sufficient to allow DNA synthesis to continue until terminated by the agent, e.g., a ddNTP.

For cycle sequencing, the DNA polymerases of the present invention make it possible to use significantly 20 lower amounts of dideoxynucleotides compared to naturally occurring enzymes. That is, the method involves providing an excess amount of deoxynucleotides to all four dideoxynucleotides in a cycle sequencing reaction, and performing the cycle sequencing reaction.

25 Preferably, more than 2, 5, 10 or even 100 fold excess of a dNTP is provided to the corresponding ddNTP.

In a related aspect, the invention features a kit or solution for DNA sequencing including a DNA polymerase of the present invention and a reagent 30 necessary for the sequencing such as dITP, deaza GTP, a

chain terminating agent such as a ddNTP, and a manganese-containing solution or powder and optionally a pyrophosphatase.

In another aspect, the invention features a method 5 for providing a DNA polymerase of the present invention by providing a nucleic acid sequence encoding the modified DNA polymerase, expressing the nucleic acid within a host cell, and purifying the DNA polymerase from the host cell.

10 In another related aspect, the invention features a method for sequencing a strand of DNA essentially as described above with one or more (preferably 2, 3 or 4) deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and a first chain terminating 15 agent. The DNA polymerase causes the primer to be elongated to form a first series of first DNA products differing in the length of the elongated primer, each first DNA product having a chain terminating agent at its elongated end, and the number of molecules of each 20 first DNA products being approximately the same for substantially all DNA products differing in length by no more than 20 bases. The method also features providing a second chain terminating agent in the hybridized mixture at a concentration different from the first 25 chain terminating agent, wherein the DNA polymerase causes production of a second series of second DNA products differing in the length of the elongated primer, with each second DNA product having the second chain terminating agent at its elongated end. The 30 number of molecules of each second DNA product is

approximately the same for substantially all second DNA products differing in length from each other by from 1 to 20 bases, and is distinctly different from the number of molecules of all the first DNA products having a 5 length differing by no more than 20 bases from that of said second DNA products.

In preferred embodiments, three or four such chain terminating agents can be used to make different products and the sequence reaction is provided with a 10 magnesium ion, or even a manganese or iron ion (e.g., at a concentration between 0.05 and 100 mM, preferably between 1 and 10 mM); and the DNA products are separated according to molecular weight in four or less lanes of a gel.

15 In another related aspect, the invention features a method for sequencing a nucleic acid by combining an oligonucleotide primer, a nucleic acid to be sequenced, between one and four deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and at least 20 two chain terminating agents in different amounts, under conditions favoring extension of the oligonucleotide primer to form nucleic acid fragments complementary to the nucleic acid to be sequenced. For example, the chain terminating agent may be a dideoxynucleotide 25 terminator for adenine, guanine, cytosine or thymine. The method further includes separating the nucleic acid fragments by size and determining the nucleic acid sequence. The agents are differentiated from each other by intensity of a label in the primer extension 30 products.

While it is common to use gel electrophoresis to separate DNA products of a DNA sequencing reaction, those in the art will recognize that other methods may also be used. Thus, it is possible to detect each of 5 the different fragments using procedures such as time of flight mass spectrometry, electron microscopy, and single molecule detection methods.

The invention also features an automated DNA sequencing apparatus having a reactor including reagents 10 which provide at least two series of DNA products formed from a single primer and a DNA strand. Each DNA product of a series differs in molecular weight and has a chain terminating agent at one end. The reagents include a DNA polymerase of the present invention. The apparatus 15 includes a separating means for separating the DNA product along one axis of the separator to form a series of bands. It also includes a band reading means for determining the position and intensity of each band after separation along the axis, and a computing means 20 that determines the DNA sequence of the DNA strand solely from the position and intensity of the bands along the axis and not from the wavelength of emission of light from any label that may be present in the separating means.

25 Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Description of the Preferred Embodiments

The drawings will first briefly be described.

Drawings

Figs 1-4 are the DNA sequences, and corresponding amino acid sequences, of FY2, FY3, and the DNA polymerases of *T. flavus* and *Thermus thermophilus*, 5 respectively. Figure 5 is the DNA sequence and corresponding amino acid sequence of FY4.

Examples

The following examples serve to illustrate the DNA polymerases of the present invention and their use in 10 sequencing.

Preparation of FY DNA Polymerases (FY2 and FY3)

## Bacterial Strains

*E. coli* strains: MV1190 [ $\Delta$ (*srl* - *recA*) 306::Tn10,  $\Delta$ (*lac*-*proAB*), *thi*, *supE*, F' (*traD36 proAB lacI<sup>q</sup> lacZ* 15  $\Delta$ M15)]; DH $\lambda$ <sup>+</sup> [*gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44, λ<sup>+</sup>*]; M5248 [ $\lambda$ (*bio275, cI857, cIII<sup>+</sup>, N<sup>+</sup>,  $\Delta$  (H1))].*

## PCR

Reaction conditions based on the procedure of 20 Barnes (91 *Proc. Nat'l. Acad. Sci.* 2216-2220, 1994) were as follows: 20mM Tricine pH8.8, 85mM KOAc, 200mM dNTPs, 10% glycerol, 5% DMSO, 0.5mM each primer, 1.5mM MgOAc, 2.5 U HotTub (Amersham Life Science Inc.) , 0.025 U DeepVent (New England Biolabs), 1-100 ng target DNA per 25 100ml reaction. Cycling conditions were 94°C 30s, 68°C 10m40s for 8 cycles; then 94°C 30s, 68°C 12m00s for 8 cycles; then 94°C 30s, 68°C 13m20s for 8 cycles; then 94°C 30s, 68°C 14m40s for 8 cycles.

*In vitro mutagenesis*

Restriction enzyme digestions, plasmid preparations, and other *in vitro* manipulations of DNA were performed using standard protocols (Sambrook et al., Molecular Cloning 2nd Ed. Cold Spring Harbor Press, 1989). PCR (see protocol above) was used to introduce a Phe to Tyr amino acid change at codon 667 of native Taq DNA polymerase (which is codon 396 of FY2).

Oligonucleotide primer 1 dGCTTGGGCAGAGGATCCGCCGGG (SEQ. ID. NO. 3) spans nucleotides 954 to 976 of the coding region of SEQ. ID. NO. 1 including a BamHI restriction site. Mutagenic oligo primer 2

dGGGATGGCTAGCTCCTGGGAGAGGCGGTGGCCGACATGCCGTAGA  
GGACCCCGTAGTTGATGG (SEQ. ID. NO. 4) spans nucleotides 1178 to 1241 including an NheI site and codon 396 of Sequence ID. NO. 1. A clone of exo- Taq deleted for the first 235 amino acids, pWB253 encoding DeltaTaq polymerase (Barnes, 112 Gene 29-35, 1992) was used as template DNA. Any clone of Taq polymerase or genomic DNA from *Thermus aquaticus* could also be utilized to amplify the identical PCR product. The PCR product was digested with BamHI and NheI, and this fragment was ligated to BamHI/NheI digested pWB253 plasmid to replace the corresponding fragment to create pWB253Y, encoding polymerase FY1. Cells of *E. coli* strain MV1190 were used for transformation and induction of protein expression, although any host strain carrying a lac repressor could be substituted. DNA sequencing verified the Phe to Tyr change in the coding region.

PCR primer 3 dGGAATTCCATATGGACGATCTGAAGCTCTCC (SEQ. ID. NO. 5) spanning the start codon and containing restriction enzyme sites, was used with PCR primer 4 dGGGGTACCAAGCTTCACTCCTTGGCGGAGAG (SEQ. ID. NO. 6) 5 containing restriction sites and spanning the stop codon (codon 562 of Sequence ID. NO. 1). A methionine start codon and restriction enzyme recognition sequences were added to PCR primer 5 dGGAATTCCATATGCTGGAGAGGCTTGAGTTT (SEQ. ID. NO. 7), which was used with primer 4 above.

10 PCR was performed using the above primer pairs, and plasmid pWB253Y as template. The PCR products were digested with restriction enzymes NdeI and KpnI and ligated to NdeI/KpnI digested vector pRE2 (Reddi et al., 17 Nucleic Acids Research 10,473-10,488, 1989) to make

15 plasmids pRE236Y, encoding FY1 polymerase, and pRE273Y encoding FY2 polymerase, respectively. Cells of *E. coli* strain DH $\lambda$ \* were used for primary transformation with this and all subsequent pRE2 constructions, and strain M5248 ( $\lambda$ cI857) was used for protein expression, although 20 any comparable pair of *E. coli* strains carrying the cI\* and cI857 alleles could be utilized. Alternatively, any rec\* cI\* strain could be induced by chemical agents such as nalidixic acid to produce the polymerase. The sequences of both genes were verified. pRE273Y was found 25 to produce a single polypeptide band on SDS polyacrylamide gels, unlike pRE253Y or pRE236Y.

Primer 6 dGGAATTCCATATGCTGGAACGTCCTGGAGTTGGCAGCCTC CTC (SEQ. ID. NO. 8) and primer 4 were used to make a PCR product introducing silent changes in codon usage of 30 FY2. The product was digested with NdeI/BamHI and

ligated to a pRE2 construct containing the 3' end of FY2 to create pREFY2pref, encoding FY2 DNA polymerase. Primer 7 dGGAATTCCATATGGCTCTGGAACGTCTGGAGTTGGCAGCCTCCTC (SEQ. ID. NO. 9) and primer 4 were used to make a PCR product introducing an additional alanine codon commonly occurring at the second position of highly expressed genes. The NdeI/BamHI digested fragment was used as above to create pREFY3, encoding FY3 DNA polymerase.

Preparation of FY4 DNA Polymerase

10 Bacterial Strains

*E. coli* strains: DH1 $\lambda^+$  [*gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44,  $\lambda^+$* ]; M5248 [ $\lambda$  (*bio275, cI857, cIII+, N+,  $\Delta$  (H1)*)].

PCR

15 Genomic DNA was prepared by standard techniques from *Thermus thermophilus*. The DNA polymerase gene of *Thermus thermophilus* is known to reside on a 3 kilobase AlwNI fragment. To enrich for polymerase sequences in some PCR reactions, the genomic DNA was digested prior to PCR with AlwNI, and fragments of approximately 3 kb were selected by agarose gel electrophoresis to be used as template DNA. Reaction conditions were as follows: 10mM Tris pH8.3, 50mM KCl, 800 $\mu$ M dNTPs, 0.001% gelatin, 1.0 $\mu$ M each primer, 1.5mM MgCl<sub>2</sub>, 2.5 U Tth, 0.025 U

20 DeepVent (New England Biolabs), per 100 $\mu$ l reaction. Cycling conditions were 94°C 2 min, then 35 cycles of 94°C 30s, 55°C 30s, 72°C 3 min, followed by 72 °C for 7 min.

*In vitro mutagenesis*

Restriction enzyme digestions, plasmid preparations, and other *in vitro* manipulations of DNA were performed using standard protocols (Sambrook et al., 1989). Plasmid pMR1 was constructed to encode an exonuclease-free polymerase, with optimized codons for expression in *E. coli* at the 5' end. Primer 8 (SEQ. ID. NO. 10) (GGAATTCCATATGCTGGAACGTCTGGAATTGGCAGCCTC) was used with Primer 9 (SEQ. ID. NO. 11) 5 (GGGGTACCTAACCTTGGCGGAAAGCCAGTC) to create a PCR product from *Tth* genomic DNA, which was digested with restriction enzymes NdeI and KpnI and inserted into plasmid pRE2 (Reddi et al., 1989, *Nucleic Acids Research* 17, 10473 - 10488) digested with the same enzymes. 10 To create the desired F396Y mutation, two PCR products were made from *Tth* chromosomal DNA. Primer 8 above was used in combination with Primer 10 (SEQ. ID. NO. 12) (GGGATGGCTAGCTCCTGGGAGAGCCTATGGGCGGACATGCCGTAGAGGACGCCGTAGTTCACCG) to create a portion of the 15 gene containing the F to Y amino acid change as well as a silent change to create an NheI restriction site. Primer 11 (SEQ. ID. NO. 13) (CTAGCTAGCCATCCCCCTA CGAAGAAGCGGTGGCCT) was used in combination with primer 9 above to create a portion of the gene from the 20 25 introduced NheI site to the stop codon at the 3' end of the coding sequence. The PCR product of Primers 8 and 10 was digested with NdeI and NheI, and the PCR product of Primers 9 and 11 was digested with NheI and KpnI. These were introduced into expression vector pRE2 which 30 was digested with NdeI and KpnI to produce plasmid pMR5.

In addition to the desired changes, pMR5 was found to have a spurious change introduced by PCR, which led to an amino acid substitution, K234R. Plasmid pMR8 was created to eliminate this substitution, by replacing the 5 AflIII/BamHI fragment of pMR5 for the corresponding fragment from pMR1. The FY4 polymerase encoded by plasmid pMR8 (SEQ. ID. NO. 14) is given in Figure 5.

Cells of *E. coli* strain DH1 $\lambda^+$  were used for primary transformation, and strain M5248 ( $\lambda$ cI857) was used for 10 protein expression, although any comparable pair of *E. coli* strains carrying the cI $^+$  and cI857 alleles could be utilized. Alternatively, any rec $^+$  cI $^+$  strain could be induced by chemical agents such as nalidixic acid to produce the polymerase.

15 Protein Sequencing

Determinations of amino terminal protein sequences were performed at the W.M. Keck Foundation, Biotechnology Resource Laboratory, New Haven, Connecticut.

20 Purification of Polymerases

A 1 liter culture of 2X LB (2% Bacto-Tryptone, 1% Bacto-Yeast Extract, 0.5% NaCl) + 0.2% Casamino Acids + 20 mM KPO<sub>4</sub> pH 7.5 + 50  $\mu$ g/ml Ampicillin was inoculated with a glycerol stock of the appropriate cell strain and 25 grown at 30°C with agitation until cells were in log phase (0.7-1.0 OD<sub>590</sub>). 9 liters of 2X LB + 0.2% Casamino Acids + 20 mM KPO<sub>4</sub> pH 7.5 + 0.05% Mazu Anti-foam was inoculated with 1 liter of log phase cells in 10 liter Microferm Fermentors (New Brunswick Scientific Co.). 30 Cells were grown at 30°C under 15 psi pressure, 350-450

rpm agitation, and an air flow rate of 14,000 cc/min  $\pm$ 1000 cc/min. When the OD<sub>590</sub> reached 1.5-2.0, the cultures were induced by increasing the temperature to 40-42°C for 90-120 minutes. The cultures were then 5 cooled to < 20°C and the cells harvested by centrifugation in a Sorvall RC-3B centrifuge at 5000 rpm at 4°C for 15-20 minutes. Harvested cells were stored at -80°C.

Frozen cells were broken into small pieces and 10 resuspended in pre-warmed (90-95°C) Lysis Buffer (20 mM Tris pH 8.5, 1 mM EDTA, 10 mM MgCl<sub>2</sub>, 16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% Tween 20, 0.1% Nonidet P-40, 1 mM PMSF). Resuspended cells were then heated rapidly to 80°C and incubated at 80°C for 20 minutes with constant stirring. The 15 suspension was then rapidly cooled on ice. The cell debris was removed by centrifugation using a Sorvall GSA rotor at 10,000 rpm for 20 minutes at 4°C. The NaCl concentration of the supernatant was adjusted to 300 mM. The sample was then passed through a diethylaminoethyl 20 cellulose (Whatman DE-52) column that had been previously equilibrated with Buffer A (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 300 mM NaCl, 10% glycerol, 1 mM DTT), and polymerase collected in the flow through. The sample was then diluted to a 25 concentration of NaCl of 100mM and applied to a Heparin-sepharose column. The polymerase was eluted from the column with a NaCl gradient (100-500 mM NaCl). The sample was then dialyzed against Buffer B (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 10 mM 30 KCl, 10% glycerol, 1 mM DTT) and further diluted as

needed to lower the conductivity of the sample to the conductivity of Buffer B. The sample was then applied to a diethylaminoethyl (Waters DEAE 15 HR) column and eluted with a 10-500 mM KCl gradient. The polymerase 5 was then diluted with an equal volume of Final Buffer (20 mM Tris pH 8.5, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet P-40, 100 mM KCl, 50% glycerol, 1 mM DTT) and dialyzed against Final Buffer.

#### Assay of Exonuclease Activity

10 The exonuclease assay was performed by incubating 5 ul (25-150 units) of DNA polymerase with 5 ug of labelled [<sup>3</sup>H]-pBR322 PCR fragment ( $1.6 \times 10^4$  cpm/ug DNA) in 100 ul of reaction buffer of 20 mM Tris-HCl pH 8.5, 5 mM MgCl<sub>2</sub>, 10 mM KCl, for 1 hour at 60 °C. After this time 15 interval, 200 ul of 1:1 ratio of 50 ug/ml salmon sperm DNA with 2 mM EDTA and 20% TCA with 2% sodium pyrophosphate were added into the assay aliquots. The aliquots were put on ice for 10 min and then centrifuged at 12,000g for 10 min. Acid-soluble radioactivity in 200 20 ul of the supernatant was quantitated by liquid scintillation counting. One unit of exonuclease activity was defined as the amount of enzyme that catalyzed the acid solubilization of 10 nmol of total nucleotide in 30 min at 60 °C.

25 Utility in DNA Sequencing

Example 1: DNA Sequencing with FY Polymerases (e.g., FY2 and FY3)

The following components were added to a microcentrifuge vial (0.5 ml) : 0.4 pmol M13 DNA (e.g., M13mp18, 1.0  $\mu$ g); 2  $\mu$ l Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM MgCl<sub>2</sub>); 2  $\mu$ l of labeling nucleotide mixture (1.5  $\mu$ M each of dGTP, dCTP and dTTP); 0.5  $\mu$ l (5  $\mu$ Ci) of [ $\alpha$ -<sup>32</sup>P]dATP (about 2000Ci/mmol); 1  $\mu$ l -40 primer (0.5  $\mu$ M; 0.5 pmol/ $\mu$ l 5'GTTTTCCCAGTCACGAC-3'); 2  $\mu$ l of a mixture containing 4 U/ $\mu$ l FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophosphatase (32 U/ $\mu$ l polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5  $\mu$ l. These components (the labeling reaction) were mixed and the vial was placed in a constant-temperature water bath at 45°C for 5 minutes.

Four vials were labeled A, C, G, and T, and filled with 4  $\mu$ l of the corresponding termination mix: ddA 20 termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddATP); ddT termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddTTP); ddC termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddCTP); ddG 25 termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddGTP).

The labeling reaction was divided equally among the four termination vials (4  $\mu$ l to each termination reaction vial), and tightly capped.

The four vials were placed in a constant- 30 temperature water bath at 72°C for 5 minutes. Then 4  $\mu$ l

of Stop Solution (95% Formamide 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added to each vial, and heated briefly to 70°-80°C immediately prior to loading on a sequencing gel (8% acrylamide, 8.3 M urea).

5 Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using *Taq* DNA

10 polymerase or  $\Delta$ *Taq* DNA polymerase.

Example 2: DNA Cycle Sequencing with FY Polymerases

The following components were added to a microcentrifuge vial (0.5 ml) which is suitable for insertion into a thermocycler machine (e.g., Perkin-

15 Elmer DNA Thermal Cycler): 0.05 pmol or more M13 DNA (e.g., M13mp18, 0.1  $\mu$ g), or 0.1  $\mu$ g double-stranded plasmid DNA (e.g., pUC19); 2  $\mu$ l Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM MgCl<sub>2</sub>); 1  $\mu$ l 3.0  $\mu$ M dGTP; 1  $\mu$ l 3.0  $\mu$ M dTTP; 0.5  $\mu$ l (5  $\mu$ Ci) of [ $\alpha$ -<sup>33</sup>P]dATP (about 20 2000Ci/mmol); 1  $\mu$ l -40 primer (0.5  $\mu$ M; 0.5 pmol/ $\mu$ l 5'GTTTTCCCAGTCACGAC-3'); 2  $\mu$ l of a mixture containing 4 U/ $\mu$ l FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophosphatase (32 U/ $\mu$ l polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5  $\mu$ l.

These components (labeling reaction mixture) were mixed and overlaid with 10  $\mu$ l light mineral oil (Amersham). The vial was placed in the thermocycler and 30-100 cycles (more than 60 cycles is unnecessary) from 5 45°C for 1 minute to 95°C for 0.5 minute performed. (Temperatures can be cycled from 55°-95°C, if desired). The temperatures may be adjusted if the melting temperature of the primer/template is significantly higher or lower, but these temperatures work well for 10 most primer-templates combinations. This step can be completed in about 3 minutes per cycle.

Four vials were labeled A, C, G, and T, and filled with 4 ml of the corresponding termination mix: ddA termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddATP); ddT termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddTTP); ddC termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddCTP); ddG termination mix (150  $\mu$ M each dATP, dCTP, dGTP, dTTP, 1.5  $\mu$ M ddGTP). No additional enzyme is added to the 15 20 termination vials. The enzyme carried in from the prior (labeling) step is sufficient.

The cycled labeling reaction mixture was divided equally among the four termination vials (4  $\mu$ l to each termination reaction vial), and overlaid with 10  $\mu$ l of 25 light mineral oil.

The four vials were placed in the thermocycler and 30-200 cycles (more than 60 cycles is unnecessary) performed from 95°C for 15 seconds, 55°C for 30 seconds, and 72°C for 120 seconds. This step was conveniently

completed overnight. Other times and temperatures are also effective.

Six  $\mu$ l of reaction mixture was removed (avoiding oil), 3  $\mu$ l of Stop Solution (95% Formamide 20 mM EDTA, 5 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added, and heated briefly to 70°-80°C immediately prior to loading on a sequencing gel. Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with 10 uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using Taq DNA polymerase or  $\Delta$ Tag DNA polymerase.

Example 3: Sequencing with dGTP analogs to eliminate 15 compression artifacts.

For either of the sequencing methods outlined in examples 1 and 2, 7-Deaza-2'-deoxy-GTP can be substituted for dGTP in the labeling and termination mixtures at exactly the same concentration as dGTP. When this 20 substitution is made, secondary structures on the gels are greatly reduced. Similarly, 2'-deoxyinosinetriphosphate can also be substituted for dGTP but its concentration must be 10-fold higher than the corresponding concentration of dGTP. Substitution 25 of dITP for dGTP is even more effective in eliminating compression artifacts than 7-deaza-dGTP.

Example 4: Other Sequencing methods using FY polymerases

FY polymerases have been adapted for use with many other sequencing methods, including the use of 5 fluorescent primers and fluorescent-dideoxy-terminators for sequencing with the ABI 373A DNA sequencing instrument.

Example 5: SDS-Polyacrylamide Gel Electrophoresis

Protein samples were run on a 14 X 16 mm 7.5 or 10% 10 polyacrylamide gel. (Gels were predominantly 10% Polyacrylamide using a 14 X 16 mm Hoefer apparatus. Other sizes, apparatuses, and percentage gels are acceptable. Similar results can also be obtained using the Pharmacia Phast Gel system with SDS, 8-25% gradient 15 gels. Reagent grade and ultrapure grade reagents were used.) The stacking gel consisted of 4% acrylamide (30:0.8, acrylamide: bisacrylamide), 125 mM Tris-HCl pH 6.8, 0.1% Sodium Dodecyl Sulfate (SDS). The resolving gel consisted of 7.5 or 10% acrylamide (30:0.8, 20 acrylamide: bisacrylamide), 375 mM Tris-HCl pH 8.8, 0.1% SDS. Running Buffer consisted of 25 mM Tris, 192 mM Glycine and 0.1% SDS. 1X Sample Buffer consisted of 25 mM Tris-HCl pH 6.8, 0.25% SDS, 10% Glycerol, 0.1M Dithiothreitol, 0.1% Bromophenol Blue, and 1mM EDTA. A 25 1/4 volume of 5X Sample Buffer was added to each sample. Samples were heated in sample buffer to 90-100°C for approximately 5 minutes prior to loading. A 1.5 mm thick gel was run at 50-100 mA constant current for 1-3 hours (until bromophenol blue was close to the bottom of

the gel). The gel was stained with 0.025% Coomassie Blue R250 in 50% methanol, 10% acetic acid and destained in 5% methanol, 7% acetic acid solution. A record of the gel was made by taking a photograph of the gel, by 5 drying the gel between cellulose film sheets, or by drying the gel onto filter paper under a vacuum.

Other embodiments are within the following claims.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: AMERSHAM LIFE SCIENCE

5 (ii) TITLE OF INVENTION: THERMOSTABLE DNA  
POLYMERASES

(iii) NUMBER OF SEQUENCES: 14

## (iv) CORRESPONDENCE ADDRESS:

10 (A) ADDRESSEE: Lyon & Lyon  
(B) STREET: 633 West Fifth Street  
Suite 4700  
(C) CITY: Los Angeles  
(D) STATE: California  
(E) COUNTRY: U.S.A.  
(F) ZIP: 90071-2066

## 15 (v) COMPUTER READABLE FORM:

20 (A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
storage  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: IBM P.C. DOS 5.0  
(D) SOFTWARE: Word Perfect 5.1

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: To Be Assigned  
(B) FILING DATE:  
(C) CLASSIFICATION:

## 25 (vii) PRIOR APPLICATION DATA:

Prior applications total,  
including application  
described below: one

5 (A) APPLICATION NUMBER: US 08/455,686  
(B) FILING DATE: May 31, 1995

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Warburg, Richard J.  
(B) REGISTRATION NUMBER: 32,327  
(C) REFERENCE/DOCKET NUMBER: 219/304-PCT

10 (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (213) 489-1600  
(B) TELEFAX: (213) 955-0440  
(C) TELEX: 67-3510

(2) INFORMATION FOR SEQ ID NO: 1:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1686 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ix) FEATURE:

(A) NAME/KEY: FY2  
(B) LOCATION: 1...1683

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

25 ATG CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT 48  
Met Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu  
1 5 10 15

CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA 96  
Leu Glu Ser Pro Lys Ala Leu Glu Ala Pro Trp Pro Pro Pro Glu  
20 25 30

30 GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG GCC 144

Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala  
 35 40 45

GAT CTT CTG GCC CTG GCC GCC AGG GGG GGC CGG GTC CAC CGG GCC 192  
 Asp Leu Leu Ala Leu Ala Ala Arg Gly Gly Arg Val His Arg Ala  
 5 50 55 60

CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT 240  
 Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu  
 65 70 75 80

CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC CTC 288  
 10 Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu  
 85 90 95

CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC 336  
 Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser  
 100 105 110

AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG 384  
 15 Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Glu Trp Thr  
 115 120 125

GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC AAC 432  
 Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn  
 20 130 135 140

CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG 480  
 Leu Trp Gly Arg Leu Glu Gly Glu Arg Leu Leu Trp Leu Tyr Arg  
 145 150 155 160

GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG 528  
 25 Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr  
 165 170 175

GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG 576  
 Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val  
 180 185 190

GCC GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC 624  
 30 Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly  
 195 200 205

CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT 672  
 His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe  
 35 210 215 220

GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC AAG 720  
 Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys  
 225 230 235 240

CGC TCC ACC AGC GCC GGC GTC CTG GAG GCC CTC CGC GAG. GCC CAC CCC 768  
 Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro  
 245 250 255  
 ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC 816  
 5 Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser  
 260 265 270  
 ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC 864  
 Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg  
 275 280 285  
 10 CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC AGG GGC AGG CTA AGT 912  
 Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser  
 290 295 300  
 AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG 960  
 Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly  
 15 305 310 315 320  
 CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG 1008  
 Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val  
 325 330 335  
 20 GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC 1056  
 Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser  
 340 345 350  
 GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC 1104  
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His  
 355 360 365  
 25 ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC 1152  
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp  
 370 375 380  
 CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TAC GGG GTC CTC TAC 1200  
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Tyr Gly Val Leu Tyr  
 30 385 390 395 400  
 GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG 1248  
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu  
 405 410 415  
 GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG 1296  
 35 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val  
 420 425 430  
 CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC 1344  
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr  
 435 440 445

GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC 1392  
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala  
 450 455 460  
 CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG 1440  
 5 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met  
 465 470 475 480  
 CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG 1488  
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys  
 485 490 495  
 10 CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC 1536  
 Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val  
 500 505 510  
 CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG 1584  
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val  
 15 515 520 525  
 GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG 1632  
 Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val  
 530 535 540  
 CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG 1680  
 20 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys  
 545 550 555 560  
 GAG TGA 1686  
 Glu \*

## (2) INFORMATION FOR SEQ ID NO: 2:

## 25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1689 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## 30 (ix) FEATURE:

- (A) NAME/KEY: FY3
- (B) LOCATION: 1...1686

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATG GCT CTG GAA CGT CTG GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC 48

Met Ala Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly  
 1 5 10 15

CTT CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG 96  
 Leu Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro  
 5 20 25 30

GAA GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG 144  
 Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp  
 35 40 45

GCC GAT CTT CTG GCC CTG GCC GCC AGG GGG GGC CGG GTC CAC CGG 192  
 10 Ala Asp Leu Leu Ala Leu Ala Ala Arg Gly Gly Arg Val His Arg  
 50 55 60

GCC CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG 240  
 Ala Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly  
 65 70 75 80

15 CTT CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GCC 288  
 Leu Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly  
 85 90 95

CTC CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT 336  
 20 Leu Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro  
 100 105 110

TCC AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG 384  
 Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp  
 115 120 125

ACG GAG GAG GCG GGG GAG CGG GCC CTT TCC GAG AGG CTC TTC GCC 432  
 25 Thr Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala  
 130 135 140

AAC CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC 480  
 Asn Leu Trp Gly Arg Leu Glu Gly Glu Arg Leu Leu Trp Leu Tyr  
 145 150 155 160

30 CGG GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC 528  
 Arg Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala  
 165 170 175

ACG GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG 576  
 35 Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu  
 180 185 190

GTG GCC GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC 624  
 Val Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala  
 195 200 205

GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC 672

Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu  
 210 215 220

TTT GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC 720  
 Phe Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly  
 5 225 230 235 240

AAG CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC 768  
 Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His  
 245 250 255

CCC ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG 816  
 10 Pro Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys  
 260 265 270

AGC ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC 864  
 Ser Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly  
 275 280 285

15 CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC AGC GGC AGG CTA 912  
 Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Gly Arg Leu  
 290 295 300

AGT AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT 960  
 Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu  
 20 305 310 315 320

GGG CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG 1008  
 Gly Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu  
 325 330 335

GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC 1056  
 25 Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu  
 340 345 350

TCC GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC 1104  
 Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile  
 355 360 365

30 CAC ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG 1152  
 His Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val  
 370 375 380

GAC CCC CTG ATG CGC CGG GCG AAG ACC ATC AAC TAC GGG GTC CTC 1200  
 Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Tyr Gly Val Leu  
 35 385 390 395 400

TAC GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC 1248  
 Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr  
 405 410 415

GAG GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG 1296  
 Glu Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys  
 420 425 430  
 GTG CGG GCC TGG ATT GAG ACC CTG GAG GGC AGG AGG CGG GGG 1344  
 5 Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Gly Arg Arg Arg Gly  
 435 440 445  
 TAC GTG GAG ACC CTC TTC GGC CGC CGC TAC GTG CCA GAC CTA GAG 1392  
 Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu  
 450 455 460  
 10 GCC CGG GTG AAG AGC GTG CGG GAG GCG GCG GAG CGC ATG GCC TTC AAC 1440  
 Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn  
 465 470 475 480  
 ATG CCC GTC CAG GGC ACC GCC GAC CTC ATG AAG CTG GCT ATG GTG 1488  
 Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val  
 15 485 490 495  
 AAG CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG 1536  
 Lys Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln  
 500 505 510  
 20 GTC CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC 1584  
 Val His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala  
 515 520 525  
 GTG GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC 1632  
 Val Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala  
 530 535 540  
 25 GTG CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC 1680  
 Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala  
 545 550 555 560  
 AAG GAG TGA  
 Lys Glu \* 1689

30 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

34

GCTTGGGCAG AGGATCCGCC GGG

23

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 64 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

10 GGGATGGCTA GCTCCTGGGA GAGGCGGTGG GCCGACATGC CGTAGAGGAC  
 CCCGTAGTTG ATGG

50  
64

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAATTCCAT ATGGACGATC TGAAGCTCTC C

31

(2) INFORMATION FOR SEQ ID NO: 6:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

35

GGGGTACCAA GCTTCACTCC TTGGCGGAGA G

31

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGAATTCCAT ATGCTGGAGA GGCTTGAGTT T

31

10 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 43 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGAATTCCAT ATGCTGGAAC GTCTGGAGTT TGGCAGCCTC CTC

43

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 46 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

36

GGAATTCCAT ATGGCTCTGG AACGTCTGGA GTTTGGCAGC CTCCTC

46

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGAATTCCAT ATGCTGGAAC GTCTGGAATT CGGCAGCCTC

40

10 (2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGGGTACCT AACCTTGGC GGAAAGCCAG TC

32

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 64 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 GGGATGGCTA GCTCCTGGGA GAGCCTATGG GCGGACATGC CGTAGAGGAC

50

37

GCCGTAGTTC ACCG

64

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 35 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CTAGCTAGCC ATCCCCTACG AAGAAGCGGT GGCCT

35

10 (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 1686 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: FY4  
 (B) LOCATION: 1...1683

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

20 ATG CTG GAA CGT CTG GAA TTC GGC AGC CTC CTC CAC GAG TTC GGC CTC  
 Met Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu  
 1 5 10 15

48

25 CTG GAG GCC CCC GCC CCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA  
 Leu Glu Ala Pro Ala Pro Leu Glu Ala Pro Trp Pro Pro Pro Glu  
 20 25 30

96

GGG GCC TTC GTG GGC TTC GTC CTC TCC CGC CCC GAG CCC ATG TGG GCG  
 Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp Ala  
 35 40 45

144

38

GAG CTT AAA GCC CTG GCC GCG TGC AGG GAC GGC CGG GTG CAC CGG GCA  
 Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg Ala 192  
 50 55 60  
 GCA GAC CCC TTG GCG GGG CTA AAG GAC CTC AAG GAG GTC CGG GGC CTC  
 5 Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly Leu 240  
 65 70 75 80  
 CTC GCC AAG GAC CTC GCC GTC TTG GCC TCG AGG GAG GGG CTA GAC CTC  
 Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp Leu 288  
 85 90 95  
 10 GTG CCC GGG GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCC TCC  
 Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser 336  
 100 105 110  
 AAC ACC ACC CCC GAG GGG GTG GCG CGG CGC TAC GGG GGG GAG TGG ACG  
 15 Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr 384  
 115 120 125  
 GAG GAC GCC GCC CAC CGG GCC CTC CTC TCG GAG AGG CTC CAT CGG AAC  
 Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg Asn 432  
 130 135 140  
 20 CTC CTT AAG CGC CTC GAG GGG GAG AAG CTC CTT TGG CTC TAC CAC  
 Leu Leu Lys Arg Leu Glu Gly Glu Lys Leu Leu Trp Leu Tyr His 480  
 145 150 155 160  
 GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC  
 Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr 528  
 165 170 175  
 25 GGG GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT  
 Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Leu 576  
 180 185 190  
 GCG GAG GAG ATC CGC CGC CTC GAG GAG GTC TTC CGC TTG GCG GGC  
 30 Ala Glu Glu Ile Arg Arg Leu Glu Glu Val Phe Arg Leu Ala Gly 624  
 195 200 205  
 CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTG CTC TTT  
 His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe 672  
 210 215 220  
 35 GAC GAG CTT AGG CTT CCC GCC TTG GGG AAG ACG CAA AAG ACA GGC AAG  
 Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly Lys 720  
 225 230 235 240  
 CGC TCC ACC AGC GCC GCG GTG CTG GAG GCC CTA CGG GAG GCC CAC CCC  
 Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro 768  
 245 250 255  
 40 ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC ACC AAG CTC AAG AAC 816

	Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys Asn			
	260	265	270	
	ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG AGG ACG GGC CGC			
5	Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly Arg		864	
	275	280	285	
	CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GGG AGG CTT AGT			
	Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser		912	
	290	295	300	
10	AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC			
	Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly		960	
	305	310	315	320
	CAG AGG ATC CGC CGG GCC TTC GTG GCC GAG GCG GGT TGG GCG TTG GTG			
	Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu Val		1008	
	325	330	335	
15	GCC CTG GAC TAT AGC CAG ATA GAG CTC CGC GTC CTC GCC CAC CTC TCC			
	Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser		1056	
	340	345	350	
20	GGG GAC GAA AAC CTG ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC			
	Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile His		1104	
	355	360	365	
	ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC			
	Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp		1152	
	370	375	380	
25	CCC CTG ATG CGC CGG GCG GCC AAG ACG GTG AAC TAC GGC GTC CTC TAC			
	Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Tyr Gly Val Leu Tyr		1200	
	385	390	395	400
	GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTA GCC ATC CCC TAC GAA			
	Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu		1248	
	405	410	415	
30	GAA GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG			
	Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val		1296	
	420	425	430	
35	CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC TAC			
	Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr		1344	
	435	440	445	
	GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC GAC CTC AAC GCC			
	Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala		1392	
	450	455	460	
40	CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC ATG			
	Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met		1440	

40

465

470

475

480

CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTC GCC ATG GTG AAG 1488  
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys  
 485 490 495

5 CTC TTC CCC CGC CTC CGG GAG ATG GGG GCC CGC ATG CTC CTC CAG GTC 1536  
 Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val  
 500 505 510

10 CAC GAC GAG CTC CTC CTG GAG GCC CCC CAA GCG CGG GCC GAG GAG GTG 1584  
 His Asp Glu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val  
 515 520 525

15 GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT CCC CTC GCC GTG 1632  
 Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val  
 530 535 540 545 550 555 560

GGT TAG 1686  
 Gly \*

Claims

1. An enzymatically active DNA polymerase having between 540 and 582 amino acids having a tyrosine at a position equivalent to position 667 of Taq DNA polymerase, wherein said polymerase lacks 5' to 3' exonuclease activity, and wherein said polymerase has at least 95% homology in its amino acid sequence to the DNA polymerase of Thermus aquaticus, Thermus flavus or Thermus thermophilus, and wherein said polymerase forms a single polypeptide band or an SDS polyacrylamide gel.
2. The polymerase of claim 1 wherein the amino acid sequence of said polymerase includes less than 3 conservative amino acid changes compared to one said DNA polymerase of said named Thermus species.
3. The polymerase of claim 1 wherein the amino acid sequence of said polymerase includes less than 3 additional amino acids compared to one said DNA polymerase of said named Thermus species at its N-terminus.
4. The polymerase of claim 1 selected from the group consisting of FY2, FY3 and FY4.
5. Purified nucleic acid encoding the DNA polymerase of any of claims 1-4.

6. Method for sequencing DNA comprising the step of generating chain terminated fragments from the DNA template to be sequenced with a DNA polymerase of any of claims 1-4 in the presence of at least one chain terminating agent and one or more nucleotide triphosphates, and determining the sequence of said DNA from the sizes of said fragments.

7. Kit for sequencing DNA comprising a DNA polymerase of any of claims 1-4 and a pyrophosphatase.

10 8. The kit of claim 7 wherein said pyrophosphatase is thermostable.

i< ââ?

is thermostable.

9. Apparatus for DNA sequencing having a reactor comprising a DNA polymerase of any of claims 1-4 and a band separator.

1 / 1 2

1/1	atg ctg gag agg ctt gag ttt ggc agc ctc ctc cac gag ttc ggc ctt ctg gaa agg ccc 31/11	61/21	agg gcc ctg gag ggc ccc tgg ccc 61/21
M L	E R L E F G S L H	P K A S P	L E E A P W A P
91/31	121/41	151/51	
ccg gaa ggg gcc ttc gtg ggc ttt gtt	tcc cgc aag ggg ccc atg tgg gcc gat ctt	ctg gcc gcc ggg ggc	ctg gcc ggg ggc
P E G A F V L	S R K E P M W A D	L A A R G G	L A R G G
181/61	211/71	241/81	R
gtc cac cgg gcc ccc gag cct tat aaa gcc ctc agg gac ctg aag gag ggc cgg cgg ctt	ctc agg gac ctg aag gag ggc cgg ctt	ctc gcc aaa gag ctg aag gtc gtt	ctg gcc ggg gtc
V H R A P Y K A	L R D L K E A R	L A K D S V	L A L
271/91	301/101	331/111	
agg gaa ggc ctt ggc ctc cgg ccc gag ccc atg ctc ctc gac cct tcc aac acc acc	agg ccc atg ctc ctc gac cct tcc aac acc acc	gag ggg gtc	
R E G L G P D G	D P M L A Y L	D N T P	E G V
361/121	391/131	421/141	A
cgg cgc tac ggg gag tgg acg gag ggg gag ggg gag ggg gag ctt tcc gag agg	cgg cgg gag ctt tcc gag agg ctt tcc gag agg	atc ttc gcc aac ctg tgg ggg agg ctt	atc ttc gcc aac ctg tgg ggg agg ctt
R R Y G E W T	A G E R A A	L F A N	L W G R
451/151	481/161	511/171	E
ggg gag gag agg ctc ctt tgg ctt tac egg gag gtg gag agg ccc ctt tcc gag ctc	ggg gag gag agg ctc ctt tgg ctt tac egg gag gtg gag agg ccc ctt tcc gag ctc	atg gag gcc acg ggg gtc	atg gag gcc acg ggg gtc
G E R L W L Y R	E V E R P L	L A H M	L W G R
541/181	571/191	601/201	E
gac gtg gcc tat ctc agg gcc ttg tcc ctg gag gtg gcc gag gat gcc cgc ctc gag	gac gtg gcc tat ctc agg gcc ttg tcc ctg gag gtg gcc gag gat gcc cgc ctc gag	gtc ttc cgc ctg gcc ggc cac	gtc ttc cgc ctg gcc ggc cac
D V A Y L R A	E V A E I	A H M	V A T
631/211	661/221	691/231	
ttc aac ctc aac tcc cgg gag ctc gag ctt gag ctc gag ctt gag ctc gag ctt	ttc aac ctc aac tcc cgg gag ctc gag ctt gag ctc gag ctt gag ctc gag ctt	ggc atc ggc aag aag acc ggc aag	ggc atc ggc aag aag acc ggc aag
F N L N S R D Q L E R	E R V L F D	P A I G K	V F R L A T
721/241	751/251	781/261	
cgc tcc acc agg gcc ggc gtc ctg gag gcc ctc cgc gag gcc cac atc gtg gag aag	cgc tcc acc agg gcc ggc gtc ctg gag gcc ctc cgc gag gcc cac atc gtg gag aag	atc ctg cag tac cgg gag ctc acc aag ctg	atc ctg cag tac cgg gag ctc acc aag ctg
R S T S A V L E A	H P D L P	K I Q Y R E	K I T K
811/271	841/281	871/291	
aag agc acc tac att gag ccc ttg cgc gag ctc atc cac ccc agg ggc cgc ctc cac	aag agc acc tac att gag ccc ttg cgc gag ctc atc cac ccc agg ggc cgc ctc cac	acc cgc ttc aac cag acg gcc acg	acc cgc ttc aac cag acg gcc acg
K S T Y I D P L	H P R T G	H T R F N Q	A T A T
901/301	931/311	961/321	
ggc agg cta agt agc tcc gat ccc aac ctc cag aac atc ccc gtc cgc acc ccg ctt	ggc agg cta agt agc tcc gat ccc aac ctc cag aac atc ccc gtc cgc acc ccg ctt	ggc ggg atc cgc cgg gcc ttc atc gcc gag	ggc ggg atc cgc cgg gcc ttc atc gcc gag
G R S S D P N L	P N I P V R	P R I R R	A F I A
991/331	1021/341	1051/351	
gag ggg tgg cta ttg gtc gag tat agc gag ctc egg ctg gtc gag aac ctc tcc ggc	gag ggg tgg cta ttg gtc gag tat agc gag ctc egg ctg gtc gag aac ctc tcc ggc	gac gag aac ctc ctg atc cgg gtc	gac gag aac ctc ctg atc cgg gtc
E G W L L V A	Q I E L R V	S D N L	R I V

## SUBSTITUTE SHEET (RULE 26)

Fig. 1A

2 / 12

1081/361	ttc gag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg ttc ggc gtc ccc cgg atg cgc cgg cgg F Q E G R D I H T E T A S W M F G V P R E D P L M R R A	1111/371	ttc gag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg ttc ggc gtc ccc cgg ccc ctg gac ccc cgg cgg cgg F Q E G R D I H T E T A S W M F G V P R E D P L M R R A
1171/391	1201/401	1141/381	1231/411
gcc aag acc atc aac tac tgg gtc ctc tac ggc atg tcg gcc cac cgc ctc tcc cag gag cta gcc atc cct tac gag gag gcc cag gcc A K T I N Y G V L Y G M S A H R L S Q E L A I P Y E A Q A	1291/431	ttc att gag cgc tac ttt cag agc ttc ccc aag gttc cgg ggc tgg att gag aag acc ctg gag gag ggc agg egg ggg tac gtg gag F I E R Y F Q S F P K V R A W I E K T L E K T G R R G Y E	1321/441
1261/421	1381/461	1411/471	1441/481
acc ctc ttc ggc cgc cgc tac gtt cca gag cta gag gcc cgg cgg atg cgg gtc aag aac gtc ttc ggc atg gcc ttc aac atg T L F G R R Y V P D L E A R V K S V R E A F N M	1471/491	ccc gtc cag ggc acc gcc gac ctc atg aag ctg gct atg gtt aag ctc ttc ccc agg ctg gag gaa atg ggg gcc agg atg ctc ctt P V Q G T A D M K L A M V K L F P R L E M G A R M L	1501/501
1351/451	1531/511	1591/531	1621/541
acc ctc ttc ggc cgc cgc tac gtt cca gag cta gag gcc cgg cgg atg cgg gtc aag aac gtc ttc ggc atg gcc ttc aac atg T L F G R R Y V P D L E A R V K S V R E A F N M	1561/521	ca g gtc cac gag ctg gtc ctc gag gcc cca aaa gag agg gcg gag gcc gtt gcc egg ctg gcc aag gag gtc atg ggg gtt tat Q V H D E L V L E A P K E R A E V A K E V M E G V Y	1681/561
1621/541	1651/551		P L A V P L E V G I G E D W L S A K F *

**SUBSTITUTE SHEET (RULE 26)**

Fig. 1B

1/1	atg	ctg	gaa	cgt	ctg	gag	ttt	ggc	ctc	cac	gag	ttc	ggc	ctt	ctg	gag	ggc	ccc	atg	ctg	gag	ggc	ccc	ttg	ccc
91/31	M	A	L	E	R	L	E	F	G	S	L	H	E	F	G	L	E	A	P	K	S	P	K	151/51	
	ccg	cg	gaa	ggg	gcc	ttc	gtg	ggc	ttt	gtg	ctt	tcc	cgc	aa	gag	ccc	atg	tgg	gcc	gat	ctt	ctg	gcc	ggg	ggc
181/61	P	P	E	G	A	F	V	G	F	V	L	S	R	K	E	P	W	A	D	L	A	A	R	G	G
	cgg	gtc	cac	ggg	ggc	ccc	gag	cc	cct	tat	aaa	g	A	L	R	D	G	L	K	A	D	L	S	V	L
271/91	R	V	H	R	A	P	Y	K	E	P	Y	K	E	P	D	L	K	E	A	R	G	L	A	K	
	ctg	agg	gaa	ggc	ttt	ggc	ctc	ccg	ccc	ggc	ggc	ggc	ggg	ctg	aag	gag	ggg	ctt	ctc	ggc	aaa	gac	ctg	ggc	ttt
361/121	L	R	E	G	L	G	P	G	P	G	L	A	Y	L	A	Y	L	A	Y	L	A	D	S	N	
	gcc	ggg	cgc	tac	ggc	ctc	ccg	ccc	ggc	ggc	ggc	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	
451/151	A	R	R	Y	G	G	E	W	T	E	A	G	E	R	A	A	L	S	E	R	L	F	A	N	
	gag	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg																
541/181	E	G	E	E	R	L	Y	R	E	V	E	R	P	L	S	A	V	L	A	V	A	E	T	G	
	ctg	gtc	gtc	gtc	gtc	gtc	ctt	tac	cg	gg	gtg	gag	ggg	ccc	ctt	tcc	gct	gtc							
631/211	D	V	A	Y	L	S	C	T	G	A	E	V	A	E	I	A	R	L	E	A	V	F	R	L	
	ccc	tcc	aac	ctc	aac	tcc	gg	gg	gg	gg	gg	gg	gg	gg											
721/241	P	F	N	L	S	R	D	Q	E	R	V	L	F	D	E	L	G	P	A	E	V	A	E	691/231	
	aaag	cg	cgc	tcc	acc	ggc	ggc	gtc	gtc	gtc	gtg	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	
811/271	R	S	T	S	A	A	V	L	E	A	H	P	I	V	E	K	I	Q	Y	R	E	L	T	K	
	tgt	aa	ggc	acc	acc	ttt	ggc	ctc	atc	cac	ccc	atc	gtt	gg	gg	gg	gg	gg	gg	gg	gg	gg	gg	gg	
901/301	K	R	L	S	S	S	D	P	D	L	I	H	P	R	T	G	R	L	H	T	R	F	N	961/321	
	cg	ggc	agg	ctt	agt	agg	tcc	gt	ccc	aat	atc	ccc	gtc	acc	ccg	ttc	aaac	cag	acg	gcc	atc	ccg	ggg	ggc	
91/331	E	G	W	L	L	V	A	L	D	P	Q	N	I	P	V	R	I	Q	R	I	R	A	F	I	1051/351
	ag	gag	ggg	tgg	ctt	ttg	gtg	gcc	ctg	gac	tat	agc	cag	ata	gag	gtg	ctg	gcc	cac	ctc	tcc	ggc	gac	gag	

**SUBSTITUTE SHEET (RULE 26)**

Fig. 2A

4 / 1 2

1081/361 gtc ttc cag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg ttc ggc gtc ccc 1111/371  
 V F Q E G R D I H T A S W M F V P R E A V D P L M R R  
 1171/391 gcg gcc aag acc atc aac tac ggg gtc ctc tac ggc atg tgg ccc cgc ctc tcc 1201/401  
 A A K T I N Y G V L Y G M S A H R L S Q 1231/411  
 1261/421 gcc ttc att gag cgc tac ttt cag agc ttc ccc aag gtt ggg gtc tgg att gag aag acc 1291/431  
 A F I E R Y F Q S P K V R A W I E K T 1321/441  
 1351/451 gag acc ctc ttc ggc cgc cgc tac gtt cca gag cta gag gcc cgg gtc aag agc gtt 1381/461  
 E T L F G R R Y V P D L E A R V K S V 1411/471  
 1441/481 atg ccc gtc cag ggc acc gcc gac ctc atg aag ctg gct atg gtt aag ctc ttc ccc 1471/491  
 N P V Q E T A D L M K L A M V K L F P R L E M G A R M  
 1531/511 ctt gag gtc cac gag ctg gtc ctc gag ggc cca aaaa gag ggg gtc ggg ctg gag gaa atg 1561/521  
 L Q V H D E L V L E P K A P R A V A R L A K E V M E  
 1621/541 tat ccc ctg gcc gtt ccc ctg gag gtt ggg gtt ggg ataa ggg gag gac tgg ctc tcc gcc aag 1651/551  
 Y P L A V P L E V G I G E D W L S A K E \* 1681/561  
 1141/381

Fig. 2B

SUBSTITUTE SHEET (RULE 26)

**SUBSTITUTE SHEET (RULE 26)**

## **SUBSTITUTE SHEET (RULE 26)**

Fig. 3B

7 / 1 2

acc ctc ttc ggc cgc cgg cgc tat gtg ccc gag ctc aac gcc cgg atg	2191/731	2221/741
T L F G R R Y V P D N A R V K S Y R E M A F N M	2251/751	2281/761
ccg gtc cag ggc acc gcc gac ctc atg aag ctg gcc atg gtg cgg ctt ttc	2311/771	2311/771
P V Q G T A D L M K L A M V R P R L F P R L Q E L G A R M L	2341/781	2371/791
cag gtg cac gag ctg ctc gag gcc ccc aag gac cgg gag gta gcc gct ttg	2401/801	2461/821
Q V H D E L V L E A P K D R A E R V A A K E V M E V W	2431/811	2491/831
ccc ctg cag gtg ccc ctg gtg gag gtg ggc ctg ggg gag gac tgg ctc tcc		
P Q V P L E V G L G E D W L S A K E		

SUBSTITUTE SHEET (RULE 26)

Fig. 3C

**SUBSTITUTE SHEET (RULE 26)**

Fig. 4A

9 / 12

1081/361	TTG GCG TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC CCC ATG CTC GCC TAC CTC CCC TCC ACG ACC CCC GAG	1111/371	TTG GCG TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC CCC ATG CTC GCC TAC CTC CCC TCC ACG ACC CCC GAG	1141/381	TTG GCG TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC CCC ATG CTC GCC TAC CTC CCC TCC ACG ACC CCC GAG
L A S R E G L D	L V D P M	L A Y	L D P S N	L T	P E
1171/391	GGG GIG GCG CGG CGC TAC GGG GAG TGG ACG GAG GAC GCC CAC CGG CTC CTC TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1201/401	GGG GIG GCG CGG CGC TAC GGG GAG TGG ACG GAG GAC GCC CAC CGG CTC CTC TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1231/411	GGG GIG GCG CGG CGC TAC GGG GAG TGG ACG GAG GAC GCC CAC CGG CTC CTC TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG
G V A R Y G G	T E W	A H R	L L S	L K	
1261/421	CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG	1291/431	CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG	1321/441	CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG
R L E G E E	K L W	Y H	P L	R V	A H M
1351/451	GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT GCG GAG ATC CGC CGC CTC GAG GAG GTC TTC CGC TTG GCG	1381/461	GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT GCG GAG ATC CGC CGC CTC GAG GAG GTC TTC CGC TTG GCG	1411/471	GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT GCG GAG ATC CGC CGC CTC GAG GAG GTC TTC CGC TTG GCG
V R L D V A Y	N F	A L Q	L E	R I	E V
1441/481	GGC CAC CCC TTC AAC CTC CCG GAC CAG CTG GAA AGG GTG CTC TTT GAC GAG CTT E R V L F D E	1471/491	GGC CAC CCC TTC AAC CTC CCG GAC CAG CTG GAA AGG GTG CTC TTT GAC GAG CTT E R V L F D E	1501/501	GGC CAC CCC TTC AAC CTC CCG GAC CAG CTG GAA AGG GTG CTC TTT GAC GAG CTT E R V L F D E
G H P	S R	N S	R D	R L	P A
1531/511	ACA GGC AAG CGC TCC ACC AGC GCC CGG GTG CTG GAG GCC CTA CGG GAG CCC ATC GTG GAG AGG ATC CTC CAG CAC CGG GAG CTC	1561/521	ACA GGC AAG CGC TCC ACC AGC GCC CGG GTG CTG GAG GCC CTA CGG GAG CCC ATC GTG GAG AGG ATC CTC CAG CAC CGG GAG CTC	1591/531	ACA GGC AAG CGC TCC ACC AGC GCC CGG GTG CTG GAG GCC CTA CGG GAG CCC ATC GTG GAG AGG ATC CTC CAG CAC CGG GAG CTC
T G K R S T	S A V	L E A	L R E	I V	K L Q
1621/541	ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CGG AGG ACG GGC CGC TCC AAC CAG ACG GCC	1651/551	ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CGG AGG ACG GGC CGC TCC AAC CAG ACG GCC	1681/561	ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CGG AGG ACG GGC CGC TCC AAC CAG ACG GCC
T K L K N T	Y V D P	S L V	H P R	T G	R L H
1711/571	ACG GCC ACG GGG AGG CTT AGT AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1741/581	ACG GCC ACG GGG AGG CTT AGT AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1771/591	ACG GCC ACG GGG AGG CTT AGT AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC
T A T G R L S	S D P N	S D P N	Q N I	P V R	Q R I
1801/601	GTG GCC GAG GCG GGT TGG GCG TTG GTG CTC GAC TAT AGC CAG ATA GAG CTC CGC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1831/611	GTG GCC GAG GCG GGT TGG GCG TTG GTG CTC GAC TAT AGC CAG ATA GAG CTC CGC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1861/621	GTG GCC GAG GCG GGT TGG GCG TTG GTG CTC GAC TAT AGC CAG ATA GAG CTC CGC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC
V A E A G W A	L V A	D Y S	I E L	R V	A H L
1891/631	ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG	1921/641	ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG	1951/651	ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG
I R V F Q E G	K D I	H T Q	S W M	P G	V D
1981/661	CGC CGG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTC TAC CCC TAC GAG GAG	2011/671	CGC CGG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTC TAC CCC TAC GAG GAG	2041/681	CGC CGG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTC TAC CCC TAC GAG GAG
R R A A K T	V N F	Y G M	S A H R	S Q E	L A T
2071/691	GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GGG CGG GGC	2101/701	GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GGG CGG GGC	2131/711	GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GGG CGG GGC
V A F I E R Y	F Q S	V R A W	E E K	P Y E	R K B G

**SUBSTITUTE SHEET (RULE 26)**

Fig. 4B

10 / 12

2161/721	TAC	GTG	GAA	ACC	CTC	TTC	GGG	AGA	AGG	CGC	TAC	GTG	CCC	GAC	CTC	AAC	GCC	CGG	GTG	AAG	AGC	GTC	AGG	GCG	GCG	ATG	GCC				
	Y	V	E	T	L	F	G	R	R	R	P	V	D	L	N	A	R	V	K	S	V	R	E	A	A	E	R	M	A		
2251/751	TTC	AAC	ATG	CCC	GTC	CAG	GGC	ACC	GGC	GGC	GAC	GTC	ATG	AAG	CTC	GCC	ATG	GTG	AAG	CTC	TTC	CCC	GCG	CTC	CGG	GAG	ATG	GGG	GCC	CGC	
	F	N	M	P	V	Q	G	T	A	D	L	M	K	L	A	M	V	K	M	V	P	R	L	F	P	R	E	M	G	A	R
2341/781	ATG	CTC	CTC	CAG	GTC	CAC	GAC	GAG	CTC	CTC	CTG	GAG	CCC	CTC	GCC	CCC	CAA	GCC	GGG	GCC	GGG	GCC	GCT	TTG	GCC	AAG	GAG	GCC	ATG	GAG	
	M	L	Q	V	H	D	E	L	L	L	E	A	P	Q	A	R	E	A	E	V	A	A	L	A	K	E	A	M	E		
2431/811	AAG	GCC	TAT	CCC	CTC	CCC	GTG	GAG	CTG	GAG	GTG	GAG	GGG	ATG	GGG	GAG	GAC	TGG	CTT	TCC	GCC	AAG	GGT	TAG							
	K	A	Y	P	L	A	V	E	V	E	V	E	D	V	E	D	W	L	S	A	K	G	*								

SUBSTITUTE SHEET (RULE 26)

Fig. 4C

11/12

Fig. 5A

**SUBSTITUTE SHEET (RULE 26)**

1081/361	TTC	CAG	GAG	GGG	AAG	GAC	ATC	CAC	ACC	CAG	ACC	GCA	AGC	TGG	ATG	TTC	GGC	GTC	CCC	CGG	GCG	
F	Q	E	G	K	D	I	H	T	Q	T	A	S	W	M	F	V	P	P	G	P	G	
1171/391	GCC	AAG	ACG	GTG	AAC	TAC	GGC	GTC	CTC	TAC	GGC	ATG	TCC	GCC	CAT	AGG	CTC	TCC	CAG	GAG	CTA	GCC
A	K	T	V	N	Y	G	V	L	Y	G	M	S	A	H	R	L	S	Q	E	L	A	I
1261/421	TTT	ATA	GAG	CGC	TAC	TTC	CAA	AGC	TTC	CCC	AAG	GTG	CGG	GCC	TGG	ATA	GAA	AAG	ACC	CTG	GAG	GAG
F	I	E	R	Y	F	Q	S	F	P	K	V	R	A	W	I	E	K	T	L	E	E	R
1351/451	ACC	CTC	TTC	GGA	AGA	AGG	CGC	TAC	GTG	CCC	GAC	CTC	AAC	GCC	CGG	GTG	AAG	AGC	GTC	AGG	GAG	GCC
T	L	F	G	R	R	Y	V	P	D	L	N	A	R	V	K	S	V	R	R	E	E	A
1441/481	CCC	GTC	CAG	GGC	ACC	GCC	GAC	CTC	ATG	AAG	CTC	GCC	ATG	GTG	AAG	CTC	TTC	CCC	CGC	CTC	GGG	GAG
P	V	Q	G	T	A	D	L	M	K	L	A	M	V	K	L	F	P	R	R	E	M	A
1531/511	CAG	GTC	CAC	GAC	GAG	CTC	CTG	GAG	CTC	ATG	AAG	CTC	GCC	ATG	GTG	AAG	CTC	TTC	CCC	CGC	CTC	GGG
Q	V	H	D	E	L	E	A	P	Q	A	R	A	E	V	A	A	A	A	R	M	G	A
1621/541	CCC	CTC	GCC	GTG	CCC	CTG	GAG	GTG	GAG	GTG	GGG	ATG	GGG	GAC	TGG	CTT	TCC	GCC	AAG	GGT	TAG	*
P	L	A	V	P	L	E	V	E	V	E	V	E	W	G	E	D	W	L	S	A	K	G

SUBSTITUTE SHEET (RULE 26)

Fig. 5B

## INTERNATIONAL SEARCH REPORT

Int'l Application No  
PCT/US 96/06906A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 C12N15/54 C12N9/12 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO,A,92 06188 (BARNES WAYNE M) 16 April 1992 cited in the application see the whole document ---	1-9
A	WO,A,91 09944 (CETUS CORP) 11 July 1991 see the whole document ---	1-9
A	WO,A,94 05797 (KISELEV VSEVOLOD ;SEVERIN EVGENII (RU); KORPELA TIMO (FI)) 17 March 1994 see the whole document ---	1-9
	-/-	

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

3

Date of the actual completion of the international search

1 August 1996

Date of mailing of the international search report

09.08.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

## INTERNATIONAL SEARCH REPORT

Int'l Application No  
PCT/US 96/06906

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EUR. J. BIOCHEM. (1992), 209(1), 351-5 CODEN: EJBCAI;ISSN: 0014-2956, 1992, XP000578012 RICHTER, OLIVER MATTHIAS H. ET AL: "Cloning and sequencing of the gene for the cytoplasmic inorganic pyrophosphatase from the thermoacidophilic archaebacterium Thermoplasma acidophilum" cited in the application see the whole document ---	1-9
A	WO,A,90 12111 (HARVARD COLLEGE) 18 October 1990 see the whole document ---	1-9
P,X	EP,A,0 655 506 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE, USA) 31 May 1995 cited in the application see page 6, line 15 - line 17; claims 1-52 ---	1,6,9
P,X	NATURE, vol. 376, 31 August 1995, MACMILLAN JOURNALS LTD., LONDON,UK, pages 796-797, XP002009831 M.A. REEVE AND C.W. FULLER: "A novel thermostable polymerase for DNA sequencing" see the whole document ---	1-9
P,X	AMERSHAM LIFE SCIENCE, EDITORIAL COMMENTS, vol. 22, no. 2, July 1995, pages 29-36, XP002009832 S.B. SAMOLS ET AL.: "Thermo Sequenase; a new thermostable DNA polymerase for DNA sequencing" see the whole document ---	1-9
P,A	PROC. NATL. ACAD. SCI. U. S. A. (1995), 92(14), 6339-43 CODEN: PNASA6;ISSN: 0027-8424, 3 July 1995, XP002009833 TABOR, STANLEY ET AL: "A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides" see the whole document -----	1-9

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

Int'l Application No  
PCT/US 96/06906

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9206188	16-04-92	AU-B- DE-T- EP-A- JP-T-	8906091 553264 0553264 6502303	28-04-92 28-04-94 04-08-93 17-03-94
WO-A-9109944	11-07-91	US-A- AU-B- AU-B- CA-A- EP-A- JP-T- US-A- US-A- US-A- AU-B- AU-B- AU-B- AU-B- CA-A- EP-A- JP-T- WO-A-	5322770 656315 7244491 2071213 0506889 5505105 5407800 5310652 5466591 6329694 646430 7176491 2071196 0506825 5504887 9109950	21-06-94 02-02-95 24-07-91 23-06-91 07-10-92 05-08-93 18-04-95 10-05-94 14-11-95 01-09-94 24-02-94 24-07-91 23-06-91 07-10-92 29-07-93 11-07-91
WO-A-9405797	17-03-94	FI-A- AU-B-	923911 4960893	02-03-94 29-03-94
WO-A-9012111	18-10-90	AU-B- AU-B- EP-A- JP-T- LT-A- US-A-	638246 5438290 0467953 4506002 1519 5498523	24-06-93 05-11-90 29-01-92 22-10-92 26-06-95 12-03-96
EP-A-0655506	31-05-95	AU-B- DE-U- DE-U- DE-T- ES-T- WO-A-	4193396 29513622 29513639 655506 2072238 9612042	06-05-96 19-10-95 19-10-95 28-09-95 16-07-95 25-04-96